Forthcoming in: *History and Philosophy of the Life Sciences* (2007), 29 (3): 335-358

Exploratory experimentation and scientific practice: Metagenomics and the proteorhodopsin case

Maureen A. O'Malley

Egenis, University of Exeter

Byrne House, St Germans Road

Exeter, EX4 4PJ, UK

ABSTRACT

Exploratory experimentation and high-throughput molecular biology appear to have considerable affinity for each other. Included in the latter category is metagenomics, which is the DNA-based study of diverse microbial communities from a vast range of non-laboratory environments. Metagenomics has already made numerous discoveries and these have led to reinterpretations of fundamental concepts of microbial organization, evolution and ecology. The most outstanding success story of metagenomics to date involves the discovery of a rhodopsin gene, named proteorhodopsin, in marine bacteria that were never suspected to have any photobiological capacities. A discussion of this finding and its detailed investigation illuminates the relationship between exploratory experimentation and metagenomics. Specifically, the proteorhodopsin story indicates that a dichotomous interpretation of theory-driven and exploratory experimentation is insufficient, and that an interactive understanding of these two types of experimentation can be usefully supplemented by another category, 'natural history experimentation'. Further reflection on the context of metagenomics suggests the necessity of thinking more historically about exploratory and other forms of experimentation.

Exploratory experimentation and scientific practice: Metagenomics and the proteorhodopsin case

Metagenomics is a rapidly developing molecular microbiological approach that combines high-throughput techniques with a focus on naturally occurring communities of microorganisms. For some observers, metagenomics refers merely to the sequencing and bioinformatic analysis of large amounts of DNA from whole groups of diverse microorganisms in specific environments ('environmental DNA'). From this point of view, metagenomics is different in quantity but not kind from ordinary single-taxon microbial genomics, and similarly shallow. Because it is not usually thought of as experimental in the sense of testing theoretically derived hypotheses, metagenomics is given a lower scientific status than many more traditional practices in microbiology. Closer analysis shows, however, that metagenomic research programmes are making major scientific discoveries through the application of a diverse range of methods and techniques, while simultaneously contributing to the revision of many of the most basic conceptual frameworks in microbiology. These revisions include the species concept, theories of biodiversity and biogeography, and guiding ideas about the fundamental units of microbial investigation. The metagenomic objective to investigate currently uncharacterized microbial entities and processes from multiple angles means that the description of 'exploratory experimentation' could be a highly apt one.

In what follows, I will first describe how metagenomic exploration has detected a plethora of new regularities in the microbial world and how the field's systematic but flexible approach extends earlier theoretical frameworks while sidestepping some of the practical and conceptual blockages imposed by more conventional microbiological research programmes. Second, I will show how exploratory experimentation has worked in one case of metagenomic analysis, the unexpected finding and subsequent investigation of proteorhodopsin genes in oceanic bacteria by Ed DeLong and colleagues. Finally, I will draw out the broader implications of this case and metagenomic inquiry for the notion of exploratory experimentation and how philosophers and historians of biology might further develop it.

1. Exploratory experimentation

Exploratory experimentation is a description of scientific practice that Friederich Steinle and Richard Burian arrived at separately in the 1990s (Steinle 1997; 2002; Ribe and Steinle 2002; Burian 1997; 2001). It has been more broadly discussed in a small number of other papers as experimental practice has become a focus of philosophical and historical studies of science (e.g., Waters 2004; Rheinberger 1997; Franklin 2005). Accounts of exploratory experimentation (EE) emphasize the role of systematically varied experimentation in scientific developments and claim that it is often conducted without specific theoretical tests in mind as new phenomena and processes are explored. Frequently, such work results in new conceptual frameworks

and bodies of knowledge. Steinle in particular has stressed the differences between theory-driven inquiry and exploratory experimentation.

TABLE ONE: Theory-driven versus exploratory experimentation (based on Steinle 1997)

	Theory-driven (TDE)	**Exploratory (EE)**
Aim	To test specific expectations	To obtain empirical regularities; generate concepts and classifications
Phenomena	Simple/simplified systems	Complex interacting systems
Research basis	Theoretically derived specific research question; isolated experiment	Broad inquiry based on multiple experiments and their relationships
Theory	Well formed; guiding	Often unavailable
Instrument	Specifically designed	Flexible
Historical period	Field already formed; expanding and refining	'Epistemic situations' rather than field/tradition/period
Consequence	Theory clarification	New and fundamental conceptual framework

Despite these differences, TDE and EE are also conceived of as having many features in common. Both forms of inquiry are systematic, progressive (i.e., they result in improved understanding), descriptive as well as explanatory, and evaluative of existing concepts and theories. In the following discussion, I will examine the distinctions made between theory-driven and exploration-driven scientific practice in relation to the new genomic field of metagenomics, with a particular focus on the experimental investigations that followed the metagenomic discovery of proteorhodopsin genes in marine bacteria.

2. Metagenomics

Microbiology has always been a technology dependent science, from the invention of the microscope through to the development of high-throughput sequencing technologies. Recent advances in understanding microbial diversity and ecoysystem function have been made on the basis of earlier achievements in microbial phylogeny and microbial genomics. Molecular

approaches to phylogeny have revolutionized evolutionary microbiology and microbial systematics. Conventionally, these microbial phylogenies have been constructed on the basis of analyses of sequences of ribosomal RNA genes (rDNA). Due to the essential importance of the ribosome in cellular life, ribosomal genes have been treated as pivotal units of evolutionary comparison (Woese and Fox 1977).[1] Traditionally, this DNA has been taken from well-characterized isolated organisms that have been 'proven' by pure laboratory cultures to be properly classified and non-aberrant lineages.

A 'logical extension' of sequencing the DNA of cultured organisms is to sequence environmental DNA (DeLong 2007; Pace 1997). The rationale for such a move lies in what is commonly called the 'great plate count anomaly', which is the discrepancy between counts made under a microscope of living cells in sample material such as sea water, and the much lower cell counts in the cultures that can subsequently be grown from those samples (Staley and Konopka 1985). One conclusion drawn from this persistent laboratory problem has been that uncultured and perhaps unculturable microbes are in the majority (often estimated to be 99% or more of microbial biodiversity) and that as a consequence, microbial biodiversity has eluded and may be beyond the reach of traditional methods (Amann et al. 1995). As part of the attempt to capture microbial biodiversity more systematically, microbiologists turned to genomics – a technology and mode of analysis that has arguably made many of its most powerful advances in the study of microbes (Fraser-Liggett 2005).

All the earliest whole genome sequences were viral and prokaryotic (including the ΦX174 bacteriophage, the bacteria *Haemophilus influenza* and *Mycoplasma genitalium*, and the archaeon *Methanocaldococcus jannaschii*[2]), and they functioned as 'proof in principle' that whole genome sequencing was not only technologically feasible but scientifically valuable. And today, by far the majority of whole genome and other sequences are microbial,[3] which means that the comparative and functional genomic knowledge gleaned from microbial genomes greatly exceeds that of plant, fungi and animal genomes (Binnewies et al. 2006; Fraser-Liggett 2005). As in molecular biology, microorganismal tractability made unicellular organisms a focus of much sequencing activity and the source of many technological breakthroughs as well as evolutionary and ecological insights.

Metagenomics – sometimes called 'environmental genomics' or 'community genomics' – represents the point at which interests in genome sequencing and uncultured organisms meet. It is an approach that analyses large amounts of microbial DNA taken directly from a wide range of environments. Environments of interest include any occupied niche, from oceans and

[1] Prokaryote rRNA and rDNA is generally considered comparable to eukaryote rRNA and rDNA (e.g.: Patterson and Sogin 2000), although eukaryote rRNAs are longer, the ribosomes bigger, and the genes not organized as operons (as they are in prokaryotes) but as multiple copies in tandem repeats.

[2] At the time of sequencing, *Methanocaldococcus jannaschii* was named *Methanococcus jannaschii*.

[3] See www.ncbi.nlm.nih.gov/genomes/static/gpstat.html for frequently updated statistics of completed genome projects.

swamps to human intestines and drinking water valves.[4] DNA is collected without initial discrimination in regard to individual organisms, taxa or particular genes, although filtering protocols screen out macroorganisms or other material.[5] The 'meta' prefix of metagenomics is sometimes given three interlinked interpretations: the field *transcends* culturing limitations and obstacles to understanding microbial biodiversity; it generates an *overarching* understanding of genetic diversity from a world that will never be exhaustively sampled; it aims to achieve an *aggregate*-level approach to biology, not an individual organism or single-genome focus (Committee on Metagenomics 2007).

Currently, metagenomicists take two approaches to the study of metagenomes. The most common consists of extracting DNA from environmental samples and cloning it in large-insert libraries.[6] These are screened for clone activity (particular functions expressed in the host cell) or specific gene sequences (Riesenfeld et al. 2004). High-interest genes continue to include rRNA genes, which are used as phylogenetic 'anchors' for further analysis of diversity and function (Tringe and Rubin 2005).

The second metagenomic approach is even more comprehensive and involves the random shotgun-sequencing of small-insert libraries[7] of every nucleotide in an environmental sample. 'Classic' examples of this approach include the sequencing of a biofilm consisting of just a few taxa from the highly acidic metal-rich runoff of a mine (Tyson et al., 2004) and of a considerably more complex oceanic community in the low-nutrient Sargasso Sea (Venter et al. 2004).[8] Metagenomics also leads inevitably to metatranscriptomics, metaproteomics and metametabolomics, in which whole-community gene expression, protein expression and metabolite levels are measured and analysed (Wilmes and Bond 2006; Ram et al. 2005; Gill et al. 2006).

The scientific status of metagenomics

For some biologists, and potentially many philosophers of biology, metagenomics is nothing more interesting than DNA sequencing on an even broader scale than before. While such activities may generate a lot of sequence information, sceptical and favourable commentators alike view this information as undirected and only shallowly informative, because the inquiry

[4] Sampling locations are specified by researchers but are usually believed to be representative of particular habitats.

[5] Aquatic sampling, for example, is conducted with filters sized to exclude larger organisms. In soils, a range of physical and chemical techniques remove other materials, such as organic matter (Tringe and Rubin 2005).

[6] DNA fragments of 100 kb and more can be propagated in bacterial artificial chromosomes (BACs) and up to 40 kb in fosmids (modified plasmids).

[7] These libraries consist of very small fragments of DNA cloned in plasmids and amenable to high-throughput sequencing.

[8] The Sargasso Sea metagenome used to be the biggest metagenomic dataset, but was recently superseded by the Venter team's 6.25 gigabase metagenomic inventory that was compiled from sequence collected from oceans round the world (Rusch et al. 2007).

is not driven by hypotheses derived from theories (Oremland et al. 2005; Ward, 2006). Very similar observations and complaints have been made about single-taxon genomics ('monogenomics') and still are (e.g., Kell and Oliver 2003; Brent 2000; Allen 2001). These 'fishing expeditions' are often contrasted with the more orderly and systematic nature of proper hypothesis-driven science (Lockhart and Winzeler 2000). Numerous microbiologists go further, arguing that the difficulty of reassembling individual genomes from metagenomic sequences results in such studies lacking any real biological meaning (e.g., Steward and Rappé 2007). Phenotype information is essential, they argue. This argument currently informs something of a backlash against the more ardent metagenomic claims about the majority of microbes being unculturable (rather than just uncultured), and has resulted in renewed interest in expanding methods for culturing so far recalcitrant taxa (e.g., Glausiusz 2007; Stevenson et al. 2004). Even advocates and practitioners of metagenomics acknowledge that the abundance of metagenome sequence data has completely outstripped the scientific community's ability to interpret it and derive findings about function (DeLong 2007).

Although metagenomics may still be limited in regard to the extent to which its copious data can be interpreted, it has already generated a whole new understanding of biodiversity and its distribution (Koonin 2007). Viral metagenomics, for example, is finally giving an indication of the diversity and activity of the prolific genetic reservoirs constituted by viruses and how they fundamentally shape microbial communities (Hambly and Suttle 2005; Edwards and Rohwer 2005). While the ocean metagenome is dominated by just a few genera of prokaryotes, it is now known that there is far more diversity than ever anticipated at the 'species' level, or more properly, at the level of the ribotypes or strains defined by rRNA genes (Rusch et al. 2007; Koonin 2007). The genomic heterogeneity in environmental samples shows that the genomes of particular isolates can no longer be assumed to be typical of whole populations or species (Allen and Banfield 2005). This diversity is frequently interpreted as adaptive in relation to the ecological roles each subtype plays in a defined ecosystem (Rusch et al. 2007). The scope and adaptive importance of the variability indicated by metagenomic analysis has generated huge biotechnological hopes. [9]

Metagenomic inventories and analyses are increasingly directed towards gaining dynamic understandings of the structure and dynamics of microbial consortia and ecosystems (Riesenfeld et al. 2004). An important focus of investigation is how responses to different environmental gradients result in different metabolic strategies, which may spread through communities both vertically, as adaptations inherited from parental cells, and horizontally, via various mechanisms that operate outside reproduction between one organism and another (DeLong et al. 2006). The extended understanding of biological systems that is generated by community-level molecular analyses has led

[9] Programmatic statements about the commercial applications of metagenomics focus on the discovery of 'novel natural products' such as enzymes, antibiotics and other drugs as well as more effective bioremediation techniques (Cowan et al. 2005; Pazos et al. 2003).

some metagenomicists to argue that communities should be understood as metaorganisms, and that metagenomics ultimately leads to biosphere-level understanding (e.g., Kowalchuk et al. 2007; Committee on Metagenomics 2007).

Some recent and highly illuminating biogeochemical findings in microbiology have been developed on the basis of clues offered by metagenomic data. One example is the exploration of the role of methane consumption in anoxic deep-sea sediments by archaeal methanotrophs with the help of bacterial sulphate reducers (Hallam et al. 2004). 'Reverse methanogenesis' or methane consumption is the reverse of the better studied process of methane production (methanogenesis) in archaeal communities. Microbial methane production is well known but the deep-sea sediment organisms involved in anaerobic methane oxidation (coupled with those reducing sulphate) have proved resistant to pure culture experimentation. Metagenomic study of this process has led to a hypothetical model of reverse methanogenesis and may have solved the biogeochemical puzzle of why seabed methane does not escape into ocean waters and the atmosphere (Hallam et al. 2004).

Probably the most developed and celebrated metagenomic discovery, however, is that of a whole new class of genes in the rhodopsin family, now called proteorhodopsin genes (DeLong 2005). I will focus on this example – the 'ecological posterchild of metagenomics success' (Kowalchuk et al. 2007) – to show the depth of insight metagenomics enables, as well as how this new field can be interpreted in relation to exploratory experimentation.

3. The discovery and investigation of proteorhodopsin

Proteorhodopsin (PR) genes were first discovered in 2000 in an uncultured marine gammaproteobacterium group, called SAR86, which had been sampled from Monterey Bay (Béjà et al. 2000). Previously these organisms were considered to be exclusively chemoorganotrophs (gaining energy from the oxidation of organic compounds), and non-photosynthetic phototrophy[10] was not suspected to exist in oceans. The proteorhodopsin gene was stumbled upon in a 130kb stretch of environmental DNA when the flanking regions of an identified rRNA gene were sequenced.[11] The inferred amino acid sequence of the flanking nucleotides was tentatively recognized as similar in structure to other photoproteins (Béjà et al. 2000). To see if the function was also similar, copies of the sequence were inserted into retinal-

[10] Phototrophy is the conversion of light energy to chemical energy via a simpler process than that which occurs in photosynthesis. Photosynthesis is a special kind of phototrophy that involves protein complexes containing chlorophyll (Bryant and Frigaard 2006). It is the standard form of energy production in cyanobacteria, such as Synechocystis and Prochlorococcus.

[11] Sequencing DNA around the rRNA gene is a common enough practice – sometimes called 'phylogenetically anchored chromosome walking' (DeLong 2005: 461) – and is done for convenience rather than to test any specific hypothesis about the functionality of sequence near rDNA.

enhanced laboratory *E. coli* in the hope that PR would be expressed and the activity of light-driven energy production observed. It was. This achievement was another bit of luck because other microbial rhodopsins (on which more below) were known not to express well in *E. coli* (Delong 2005: 461).

The PR discovery startled the whole marine microbiology community and galvanized numerous PR-focused surveys of marine DNA samples. A huge variety of PRs in SAR86 and other bacterial groups were detected by subsequent metagenomic and genetic analyses (de la Torre et al. 2003; Sabehi et al. 2004). The Sargasso Sea metagenome alone contains 800 rhodopsin homologues (Venter et al. 2004) and recent research has even found freshwater bacteria that possess proteorhodopsin variants (Sharma et al. 2008). Retinal-binding membrane pigments or proteorhodopsins are now known to exist in 13% of marine microorganisms, many of which are gammaproteobacteria and alphaproteobacteria (Sabehi et al. 2005). There are an estimated 10^{28} PR-expressing bacterial cells in the Earth's oceans, and their abundance means they are amongst the most prolific and presumably successful organisms on the Earth (Morris et al. 2002).

In prokaryotes, rhodopsin was previously known only in extremely halophilic or salt-loving archaea (haloarchaea), where it was called bacteriorhodopsin (BR), despite being archaeal and not normally bacterial.[12] BR was discovered in the early 1970s in the halophilic archaeon, *Halobacterium salinarum*[13] – a 'startling conclusion' for the researchers involved (Oesterhelt and Stoeckenius 1971; 1973) because light-driven energy production that was not chlorophyll-based had never been anticipated.[14] The name BR was initially deemed a poor choice, because little was known about visual rhodopsin and significant similarity between it and BR was not expected (Henderson and Schertler 1990; see Table Two for a technical comparison of the two rhodopsin types).

[12] BR was discovered in the 1970s, and the domain (superkingdom) level division between bacteria and archaea was not established until the 1980s. Major differences between archaea and bacteria include the chemical structure of their cell membranes and their transcriptional and translational machinery.

[13] Halobacteria were named prior to the discovery of archaea, and are commonly called haloarchaea now. *Halobacterium salinarum* has had several other species names, including *H. halobium* (its name when rhodopsin was discovered in its membrane).

[14] I and other colleagues have recently embarked on a project in which we will compare the BR story to the PR one in order to gain a richer understanding of the activities in each discovery narrative.

TABLE TWO: Rhodopsin superfamilies

RHODOPSINS:

Retinylidene proteins in membranes, characterized by seven transmembrane alpha-helices connected by interhelical loops. They form a pocket binding the chromophore (the light absorbing unit), which is a covalent carotenoid retinal (Vitamin A aldehyde).

Type I
Microbial (archaeal, eukaryotic, bacterial)

1. **Archaeal**: transporters and sensory
a) Transporters
i) *Bacteriorhodopsin* (BR)
Photoactivation of retinal expels protons into the periplasm and creates a proton gradient that drives ATP production and other energy demands
ii) *Halorhodopsin* (HR)
Light-driven transport of extracellular chloride ions into the cytoplasm increases the electrochemical potential of the proton gradient
b) Sensory receptors
i) *SRI*
Low oxygen levels lead to expression of SRI, which directs phototaxis in a positive response to orange light (optimal for BR and HR) and negative to UV light
ii) *SRII*
High oxygen levels lead to expression of SRII and avoidance of blue light for protection from photoxidative damage

2. **Eukaryotic**: primarily sensory
Proton pump in one fungus
Sensory rhodopsins identified bioinformatically in algae, yeasts and fungi

3. **Bacterial** (**proteorhodopsin**: PR)
a) Proton pump: same function as BR
b) Sensory receptor genes in marine bacteria identified bioinformatically in association with PR genes

Type II
Animal photoreceptors in eyes and other tissues

Isomerization of retinal, the primary vision event, induces photosensory signals through retinal interactions with other proteins

Type II and Type I comparison

Similarities
- Seven transmembrane alpha-helices form a binding pocket for retinal
- Retinal is similarly attached in the seventh helix
- Retinal photoisomerization results in similar photochemical reactions

Differences
- Different conformational changes of retinal
- Little sequence similarity
- Seven helices arranged differently
- BR, HR and PR function very differently after photoisomerization than do visual rhodopsins
- Sensory receptors interact with different proteins

BR's amenability to multiple experimental techniques and its positive implications for Mitchell's then controversial chemiosmotic theory combined to make it a model object of study in relation to membrane proteins and transport mechanisms (Oesterhelt and Stockhenius 1973; Lanyi 2004). A great deal is now known about the general structure and function of rhodopsins due to a wide range of experimental analyses first conducted with BR (Spudich 2006; Lanyi 2004; see Table Two). As well as providing a mechanism for using light energy, BR is often coupled to sensory rhodopsins, which enable the detection of light wavelengths and movement to or away from beneficial or harmful light (see Table Two). Over 30 BR homologues have since been found in other archaea, as well as in microbial fungi, algae and yeast (Spudich et al. 2000; Waschuk et al. 2005).[15]

The early findings that PR functioned as a light-driven proton pump that increased ATP production were rapidly backed up by numerous structural and functional analyses employing biophysical, biochemical and genetic techniques (Martinez et al. 2007). Combinations of sequencing and experimental analysis have shown that distinct light absorption spectra are associated with rhodopsin variants, which are spectrally tuned to light availability in shallow or deeper water (Béjà et al. 2001; Man et al. 2003; Walter et al. 2007). Other studies, primarily bioinformatic but also experimental, have tentatively identified sensory receptor genes in association with PR genes (Spudich 2006; Wang et al. 2003). Initial physiological experiments were unable to detect any fitness advantages to PR^+ organisms (Giovannoni et al. 2005), but subsequent research has shown that light has a positive impact on growth in PR^+ organisms in low-nutrient, respiratorily challenging conditions (Gómez-Consarnau et al. 2007; Walter et al. 2007).

Increasingly broad phylogenetic studies of PR genes show that they are patchily distributed across a wide range of bacterial lineages, and that they have probably been transferred across lineages numerous times (Sharma et al. 2006). Interesting phylogenetic patterns have spurred further PR searches with the aim of generating a clearer evolutionary picture. Recently, a new rhodopsin has been found in the halophilic bacterium, *Salinibacter ruber*. Xanthorhodopsin, the previously unknown rhodopsin, shares some light-harvesting similarities with chlorophyll-based complexes in plants and cyanobacteria (Mongodin et al. 2005; Balashov et al. 2005). Overall, however, *S. ruber* genes group phylogenetically with archaeal rhodopsins (indicating transfer from archaea to bacteria), whereas deep-sea archaeal rhodopsins appear to be acquisitions from bacteria (Sharma et al. 2006; Frigaard et al. 2006). Such findings raise pressing questions about the evolutionary relationships of the various rhodopsins, and also inspire further sequencing forays into microbial metagenomes.

[15] Possibly also in cyanobacteria (Jung et al. 2003), although the phylogenetic relationship of the recently discovered sensory receptor rhodopsin in *Anabaena* to archaeal rhodopsins is not clear (Sharma et al. 2006).

4. Proteorhodopsin findings and theory

The rapidly expanding body of PR research is leading to novel theoretical formulations as well as challenges to old theoretical frameworks. The extensive distribution and diversity of proteorhodopsin genes was one of the first findings gained in metagenomic studies of PR. As a cascade of experimentation using a plethora of tools and techniques followed this early work, new pictures began to emerge of community-level dynamics and the role of rhodopsins in complex ocean ecologies (de la Torre et al. 2003), including further unanticipated links to sulphur cycling (Sabehi et al. 2005). PR studies have, in fact, provided a completely new picture of the nature and prevalence of light utilization in ocean waters (DeLong 2005).

PR research has also developed the notion of the 'habitat genome', which is a pool of genes useful for adaptation to particular environments (Mongodin et al. 2005). 'Cosmopolitan genes', such as PR genes, which have undergone extensive lateral gene transfer and positive selection, indicate that what is important to study are functional properties in relation to particular environments, rather than specific organisms or organismal lineages (Frigaard et al. 2006). Metagenomicists have found it valuable to focus on the laterally transferred mobile genetic elements[16] that are implicated in the spread of the PR photosystem, in order to understand what functions are most ecologically in demand. PR genes are therefore considered to be 'niche defining' (Kowalchuk et al. 2007: 479), in that their sequences say a great deal about the geochemical circumstances in which they occur. Further knowledge about PR function may lead to new understandings of geochemical cycles, such as carbon flow in oceans, especially if switching on phototrophy also means switching on as-yet undetected abilities of the relevant organisms to fix CO_2 (Giovannoni and Stingl 2005; Bielawski et al. 2004; Béjà et al. 2000).

Numerous evolutionary hypotheses are being generated from and analysed in relation to the quantities of rhodopsin data that have come out of metagenomic and associated projects. The lateral transfer of rhodopsin genes appears to have occurred frequently within domains (superkingdoms) and even across domains, between archaea and bacteria (PR-affilated genes have been found in archaea and BRs in the bacteria). Although photosynthesis is a very efficient means of light harvesting, photosynthesis genes tend not to spread laterally because of the large gene complex (~30 genes) they comprise (Bryant and Frigaard 2006; McCarren and DeLong 2007). Microbial rhodopsins underpin much less efficient light-based energy systems, and their ubiquity and frequent transfer are probably due to the fact that rhodopsin gene complexes are much smaller than photosynthesis gene

[16] Mobile genetic elements are sequence fragments able to relocate themselves in other genomes. They include plasmids, gene cassettes, bacteriophages, transposons and genomic islands (Frost et al. 2005; Binnewies et al. 2006). Mechanisms of transfer include conjugation (the microbial equivalent of mating), transduction (the phage-mediated transfer of DNA) and transformation (when cells switch on capabilities to pick up and integrate free DNA from the environment). Transfer has to be followed by integration into the genome, through legitimate or illegitimate recombination (Thomas and Nielsen 2005).

complexes (only six genes are required even in naturally non-phototropic organisms, such as the *E. coli* used as host cells). They can therefore easily be carried unexpressed in genomes until stressful conditions occur. In addition, microbial rhodopsins form highly plastic photosystems, compatible with diverse organisms from many taxa and with different cell membranes (Martinez et al. 2007). Frequent transfer occurs, therefore, because the acquisition of phototrophic capacities is of low cost and high benefit (McCarran and DeLong 2007; Sharma et al. 2006; Martinez et al. 2007; Frigaard et al. 2006).

It seems likely that proteorhodopsins – far more widespread than archaeal rhodopsins – are the ancestral rhodopsins (Mongodin et al. 2005). Microbial sensory rhodopsins are younger in evolutionary terms than light-driven proton pumps, but may have evolved before photosynthesis (Spudich 2006). Proliferating findings of a variety of PR and BR genes and their transfer across organismal superkingdoms raise questions about whether there was an earliest common ancestor that possessed BR/PR before the divergence of the three domains or whether there was a single bacterial origin followed by LGT and diversification (Ruiz-González and Márin 2004).

There is speculation that further metagenomic connections will be found between photosynthesis and rhodopsin-based phototrophy, which are currently believed to have evolved independently from different origins (Bryant and Frigaard 2006). Making broader evolutionary links between apparently analogous systems also stimulates questions about whether Type I (microbial) and Type II (visual) rhodopsins might not, after all, have a common evolutionary origin. Although there is little sequence homology between visual and microbial rhodopsins, protein architecture and general photochemical function are similar (see Table Two). Some researchers anticipate that further metagenomic data might disclose organisms possessing genes for both types, but currently think of the evolutionary relationship as one of convergence (Ruiz-González and Márin 2004; Sharma et al. 2006). More radical speculations prompted by the discovery of proteorhodopsin include the proposal of a common origin for all rhodopsins (Gehring 2004; for several reasons Gehring is likely to be wrong, see Fernald 2006). It is clear that explorations of rhodopsin function in bacteria, and of the evolutionary relationships between PR and other rhodopsins, are stimulating fundamental reconsiderations of existing evolutionary accounts.

5. Implications for exploratory experimentation

Proteorhodopsin was an accidental discovery of metagenomic analysis. One of the key marine microbiologists involved, Oded Béjà, describes the initiation of PR inquiry as 'flying blind' and a matter of luck (Béjà, in Sreenivasan 2001). Marine DNA was sequenced because it was there, and putative rhodopsin-encoding sequence analysed with nothing more theoretical in mind than the general idea it could be similar to bacteriorhodopsin. Prima facie, it is hard to deny the highly exploratory nature of both the original discovery and the subsequent barrage of experimental techniques and analyses to which

proteorhodopsin was subjected. How does this research measure up against Steinle's and Burian's more specific criteria of exploratory experimentation?

'Systematic variation of experimental parameters', the first criterion outlined by Steinle and Burian, is a key characteristic, and it would seem the proteorhodopsin story meets this criterion in a broad sense (if we do not think of experimentation as being restricted to specific indoor laboratory situations). Variation of conditions was sought by altering geographical location (different oceans and latitudes), environmental conditions (ocean depth and light availability, oxygen levels, nutrients, season and temperature), comparative level (superkingdoms, species, populations, species, strains or ecotypes), and – of course – investigative technique (biophysical, biochemical, genetic, bioinformatic, phylogenetic).

'New regularities and empirical patterns' (Criterion Two) are certainly the outcome of all this multiperspectival study, and new concepts and even more fundamental conceptual frameworks (Criterion Three) are already, in the seven short years in which proteorhodopsin has been studied, developing in relation to photobiology, ocean ecology, and rhodopsin evolution.

On the basis of meeting these criteria (more systematically discussed in Kevin Elliott's paper), the proteorhodopsin case would seem to fall under the rubric of exploratory experimentation. The very fact it does, however, generates other questions about the fit of this case to EE, and these have implications for current ideas about exploratory experimentation. The first issue, and one that Franklin (2005) has also observed, is that background knowledge and general concepts frame even the most basic high-throughput data gathering. The PR discovery could only have been made in the context of knowledge about microbial rhodopsins in general, and the stimulus of the discovery was so strong because of the potential challenges PR offered to existing understandings of microbial rhodopsin. Furthermore, as new concepts and classificatory frameworks are being developed to take proteorhodopsin into account, older theoretical frameworks (e.g., about rhodopsin evolution) are being called into service and examined for their adequacy in relation to this new knowledge. In other words, hypothesis testing is brought in as part of the exploratory approach.

Steinle has himself acknowledged this, when he says, 'the new conceptual frameworks are necessarily compared with, and measured against, already established concepts and methodological standards' (1997: S72). This does not quite match, however, the very clear distinctions he has made between theory-driven and exploration-driven experimentation (see Table One above). Although he believes that there could be other types of experimentation, in which case TDE and EE would not be opposed categories of research process, in general the attraction of EE for philosophical accounts of scientific practice is precisely this opposition. The key question is whether this is the best way to understand experimental practice: as two mutually opposed types. Is this the way scientific experimentation actually occurs, or has discussion of EE perhaps made an ideal type out of TDE in order to make EE sound novel and attractive? Certainly, some of the earlier literature on EE

appears to take this position but it is clearly inadequate to the biological research discussed in this paper and to the research described by the other papers in this issue. How else might we consider these two experimental approaches?

A continuum of practices, with TDE at one end and EE at the other, might be a better way of describing the relationship between these two scientific approaches. The PR story appears to show *not* that a certain case fits on one particular point of that continuum, but that it moves back and forth as various groups engage in exploration and theory evaluation in relation to these newly emerging phenomena. One such example, already mentioned above, is the examination of amino acid variants in PR. Statistical analyses of amino acid sites in a wide range of PRs indicated positive selection of very small changes in sequence, and independent experimental manipulations of those same sites showed them to be responsible for the spectral tuning of PR proteins (Bielawski et al. 2004; Man-Aharonovich et al. 2004). Taken together, these studies suggested that small variations in PR proteins have been highly advantageous to bacteria in ocean environments (Bielawski et al. 2004). These fortuitous combinations of approach took very different routes of investigation, yet each supplemented and supported the other's results. These studies have been followed by further discovery-oriented work on how widespread such spectral tuning is, and whether there are differences in the spectral tuning of PRs in different oceans (Sabehi et al. 2007).

Do such examples mean that TDE and EE exist as interactive processes of inquiry? This is what the proteorhodopsin story would seem to say, with non-ideal types of each approach being part of an interplay of inquiry. Steinle's own answer is to think about the other categories of experiment that exist apart from TDE and EE, and to see what other dimensions of practice may be in operation. Kevin Elliott (this issue) and Franklin (2005) suggest that what is missing in the way TDE and EE have been framed is not a more multidimensional account of experimental practice in general, but a better account of the dimensions *within* exploratory experimentation. I will complement their work by taking Steinle's suggestion to consider the dimensions of scientific practice that may occur in combination with EE and TDE.

The proteorhodopsin story suggests that among the missing dimensions of inquiry may be one that emphasizes discovery or a natural history approach, probably combined with 'natural experimentation' in which the less controllable nature of the phenomena means that strict hypothesis testing is practically impossible. This category, which we could call natural history/experimentation (NHE),[17] involves various activities of discovery, classification, comparison and probing for specific attributes or properties.

[17] Thanks to Staffan Müller-Wille for this suggestion. See Rheinberger and McLaughlin (1984) for an illuminating discussion of this category of investigation in relation to Darwin's research. William Dawson (1988) outlines 'experimental natural history' in relation to field observations of animal behaviour, for which functional capabilities can only be understood within a natural context (p. 1186).

NHE confers the status of experimenter on nature itself, and reads the results of those experiments as if they had been controlled in biologically meaningful ways (Rheinberger and McLaughlin 1984: 359). More controlled laboratory experiments can, in fact, simply be seen as idealized forms of nature's own experiments. Certain parameters are interpreted as set by nature, and these conditions are taken into account for the systematic comparison of observations.

The NHE mode of inquiry is particularly appropriate for environmental microbiology and its focus on the uncultured majority of biodiversity, the function of any part of which needs to be understood in an appropriate context of community behaviour, evolutionary relationships and geographic distribution. Very little of that context will ever be reproduced in a human-controlled laboratory setting. We can see how NHE occurs repeatedly in the PR story, when, for example, observations are made of the effects of ocean depth and light penetration on PR structure and function, or when data are collected on the effects of different ocean locations on PR distribution in different evolutionary lineages.

It is important to keep in mind, however, that these three strategies of inquiry (NHE, EE, TDE) do not occur in a serial manner from more observational to more theoretical. In the PR story, conventional experimentation, such as the manipulations of amino acid sites discussed above, triggers more discovery work (to find different absorption spectra at different depths) and combines with statistical confirmation of the adaptiveness of PR (i.e., because the amino acid variations appear to have undergone positive selection), which in turn gives rise to more experimentation and further classification work.

An alternative conceptualization of the methodology and epistemology scientific investigation that may be relevant to the PR story is offered by Robert Brandon (1994; see also Burian 1994). Concerned to avoid crude dichotomies of theory and experiment, Brandon proposes instead that a continuum of hypothesis testing interacts with another continuum of manipulation and thereby creates a space of experimentality. Degrees of manipulation range over the measurement of key variables, and degrees of hypothesis testing can decline into predominantly descriptive work. Brandon's account of practice, although proposed for evolutionary biology, could also be used to capture the interactive and non-serial ways in which different modes of experimentation are employed in any wide-ranging biological inquiry. PR investigations range from the highly descriptive to highly manipulative (e.g., from sequencing amino acids to systematically investigating what changing one amino acid residue does), and from those aimed at strict hypothesis testing to those concerned exclusively with measuring variation (e.g., from expressing PR in a laboratory organism to identifying and quantifying PR variants in a particular ocean).

However, an advantage of conceiving of PR and other biological research as interactive processes of EE, TDE and NHE means that there is room for a richer account of the dimensions of experimentality, and that extrapolations from nature's experiments to human-designed experiments can be made

more effectively. Moreover, making systematically varied observations in natural environments is more or less mandated by the inability of researchers to culture most microorganisms, and the metagenomic mode of exploratory and natural history experimentation is not adequately accounted for by Brandon's intersecting continua. In addition, his framework still describes non-manipulative non-hypothesis-testing activities as less 'powerful' science (Brandon 1994: 67). The PR story indicates that such a schema fails to capture not only some of the most remarkable moments of scientific advances, but also, their deeply systematic orientation to interpreting the results of nature's experiments. Further detailed historical work on these interacting processes of inquiry, in relation to the proteorhodopsin case and other research, would be very valuable for the philosophical and historical appreciation of experimentation in relation to biological practice in general.

Historical contextualization of exploratory science

Another factor that will be important to address in future in-depth work on EE is historical reflection on broader scientific trends and their contexts. Biology has generally placed a low value on theoretically undirected data gathering (and philosophy of science has tended to be complicit in that evaluation, as many advocates of EE mention). Jo Handelsman is one of the early metagenomicists responsible for coining the name for the new field and its research object, the metagenome (Handelsman et al. 1998; Handelsman 2004). She points out that federal and other competitive funding programmes 'reward hypothesis-driven, rather than exploratory, research' (Handelsman 2005). Early genomics was heavily criticized for its atheoretical mindlessness (Lewin 1986), and even now, the main driving force of the burgeoning 'postgenomic' field called systems biology is to bring 'omic' biology back into the fold of hypothesis-testing science (Kitano 2002).[18] Despite this resistance, scientific standards may have been changed by genomics, argues molecular biologist, Roger Brent (2000), with the result that hypothesis testing is no longer enshrined as 'the' one and only scientific method (see also Aebersold et al. 2000).

In a similar vein, philosopher Laura Franklin thinks hypothesis testing may be a reasonable account of some low-data science, but that its classic formulation is inadequate for high-throughput multidisciplinary activities (2005) – the form of scientific enquiry that is increasingly the norm in today's laboratories in academia and industry. Another factor that must be considered in contextualizing NHE, EE and TDE is the difficulty faced in publishing a scientific paper that does not conform to the TDE mode of inquiry (Sharma and Doolittle, personal communication). Whatever the actual process of inquiry and however exploratory it was, scientists often have little choice but to present it post hoc as strict hypothesis testing.

[18] See also recent National Science Foundation (US) calls for 'Advancing Theory in Biology' (www.nsf.gov/funding/pgm_summ.jsp?pims_id=501066) and the US National Academy of Sciences (2007) report, 'The role of theory in advancing 21st century biology'.

Conclusion

Rather than considering EE as having overwhelmed TDE experimentation in a particular era of biological science, it is probably worth examining more carefully exactly how alleged TDE cases really occurred. It is possible that theory-driven hypothesis testing has been conceived of by scientists, science funders and philosophers in a way that does not exist in practice (and never has), and that it is closer to and involves more interplay with exploratory experimentation as well as natural history experimentation than we have tended to think. A more dimensional, less radically conceived account of EE certainly seems to cover the metagenomics and proteorhodopsin stories – a finding that suggests a fruitful future research agenda for philosophical, historical and sociological studies of science.

Acknowledgements

I owe many thanks to Kevin Elliott for organizing the ISHPSSB session that generated this paper, and to Kevin, Dick Burian, Ken Waters, Staffan Müller-Wille and Alex Powell for comments on that paper and subsequent versions. I am also grateful to an anonymous reviewer for very helpful revision suggestions. The research for this paper was funded by the UK Arts and Humanities Research Council and was conducted under the auspices of Egenis, the Economic and Social Research Council's Centre for Genomics in Society, at the University of Exeter, UK.

References

Aebersold R., Hood, L.E., Watts J.D., 2000, 'Equipping scientists for the new biology', *Nature Biotechnology*, 18: 359.

Allen E.E., Banfield J.F., 2005, 'Community genomics in microbial ecology and evolution', *Nature Reviews Microbiology*, 3: 489-498.

Allen J.F., 2001, 'Bioinformatics and discovery: induction beckons again', *BioEssays*, 23: 104-107.

Amann R.I., Ludwig W., Schleifer K.-H., 1995, 'Phylogenetic identification and in situ detection of individual microbial cells without cultivation', *Microbial Reviews*, 59, 143-69.

Balashov S.P., Imasheva E.S., Boichenko V.A., et al., 2005, 'Xanthorhodopsin: a proton pump with a light-harvesting carotenoid antenna', *Science*, 309: 2061-2064.

Béjà O., Aravind L, Koonin E. V., et al., 2000, 'Bacterial rhodopsin : evidence for a new type of phototrophy in the sea', *Science*, 289: 1902-1906.

Béjà O., Spudich E.N., Spudich J.L., et al., 2001, 'Proteorhodopsin phototrophy in the ocean', *Nature*, 411: 786-789.

Bielawski J.P., Dunn K.A., Sabehi G., Béjà O., 2004, 'Darwinian adaptation of proteorhodopsin to different light intensities in the marine environment', *Proceedings of the National Academy of Sciences*, 101: 14824-14829.

Binnewies T.T., Motro Y., Hallin P..F. et al., 2006, 'Ten years of bacterial genome sequencing: Comparative-genomics-based discoveries', *Functional and Integrative Genomics*, 6: 165-185.

Brandon R.N., 1994, 'Theory and experiment in evolutionary biology', *Synthese*, 99: 59-73.

Brent R., 2000, 'Genomic biology', *Cell*, 100: 169-183.

Bryant D.A., Frigaard N.-U., 2006, 'Prokaryotic photosynthesis and phototrophy illuminated', *Trends in Microbiology*, 14: 488-496.

Burian R.M., 1994, 'Comments on Robert Brandon's 'Theory and experiment in evolutionary biology'', *Synthese*, 99: 75-86.

Burian R.M., 1997, 'Exploratory experimentation and the role of histochemical techniques in the work of Jean Brachet, 1948-1952', *History and Philosophy of the Life Sciences*, 19: 27-45.

Burian R.M., 2001, 'The dilemma of case studies resolved: the virtues of using case studies in the history and philosophy of science', *Perspectives on Science*, 9: 383-404.

Committee on Defining and Advancing the Conceptual Basis of Biological Sciences in the 21st Century, National Research Council, 2007, *The role of theory in advancing 21st century biology : Catalyzing transformative research*, Washington, DC: National Academies Press. Available at http://books.nap.edu/catalog/12026.html

Committee on Metagenomics: Challenges and Functional Applications, National Research Council, 2007, *The new science of metagenomics: revealing the secrets of our microbial planet*, Washington, DC: National Academies Press.

Cowan D., Meyer Q., Stafford W., et al., 2005, 'Metagenomic gene discovery: past, present and future', *Trends in Biotechnology*, 23: 321-329.

Dawson W.R., 1988, 'An experimental natural history of some terrestrial vertebrates', *American Zoologist*, 28: 1181-1188.

DeLong E.F, 2005, 'Microbial community genomics in the ocean', *Nature Reviews Metagenomics*, 3: 459-469.

DeLong E.F., 2007, 'Sea change for metagenomics?', *Nature Reviews Microbiology*, 5: 326.

DeLong E.F., Preston C.M., Mincer T., et al., 2006, 'Community genomics among stratified microbial assemblages in the ocean's interior', *Science*, 311: 496-503.

Edwards R.A., Rohwer F., 2005, 'Viral metagenomics', *Nature Reviews Microbiology*, 3: 504-510.

Fernald R.D., 2006, 'Casting a genetic light on the evolution of the eyes', *Science*, 313: 1914-1918.

Franklin L.R., 2005, 'Exploratory experiments', *Philosophy of Science*, 72: 888-899.

Fraser-Liggett, C. M., 2005, 'Insights on biology and evolution from microbial genome sequencing', *Genome Research*, 15: 1603-1610.

Frigaard N.-U., Martinez A., Mincer T. J., DeLong E.F., 2006, 'Proteorhodopsin lateral gene transfer between marine planktonic Bacteria and Archaea', *Nature*, 439: 847-850.

Frost L.S., Leplae R., Summers A.O., Toussaint A., 2005, 'Mobile genetic elements: The agents of open source evolution', *Nature Reviews Microbiology*, 3: 722-732.

Gehring W.J., 2004, 'Historical perspective on the development and evolution of eyes and photoreceptors', *International Journal of Developmental Biology*, 48: 707-717.

Gill S.R., Pop M., DeBoy R.T., et al., 2006, 'Metagenomic analysis of the human distal gut microbiome', *Science*, 312, 1355-1359.

Giovannoni S.J., Bibbs L., Cho J.-C., et al., 2005, 'Proteorhodopsin in the ubiquitous marine bacterium SAR11', *Nature*, 438: 82-85.

Giovannoni S.J., Stingl U., 2005, 'Molecular diversity and ecology of microbial plankton', *Nature*, 437: 343-348.

Glausiusz J., 2007, 'Extreme culture', *Nature*, 447: 905-906.

Gómez-Consarnau L., González J.M., Coll-Lladó M., et al., 2007, 'Light stimulates growth of proteorhodopsin-containing marine Flavobacteria', *Nature*, 445: 210-213.

Hallam S.J., Putnam N., Preston C.M., et al., 2004, 'Reverse methanogenesis: Testing the hypothesis with environmental genomics', *Science*, 305, 1457-1462.

Hambly E., Suttle C.A., 2005, 'The viriosphere, diversity, and genetic exchange within phage communities', *Current Opinion in Microbiology*, 8: 444-450.

Handelsman J., 2004, 'Metagenomics: Application of genomics to uncultured microorganisms', *Microbiology and Molecular Biology Reviews*, 68: 669-85.

Handelsman J., 2005, 'How to find new antibiotics: metagenomics could be the way to mine the soil beneath our feet', *The Scientist*, 19 (19): e20.

Handelsman J., Rondon M.R., Brady S.F., et al., 1998, 'Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products', *Chemistry and Biology*, 5: R245-49.

Henderson R., Schertler G.F.X, 1990, 'The structure of bacteriorhodopsin and its relevance to the visual opsins and other seven-helix G-protein coupled receptors', *Philosophical Transactions of the Royal Society London B*, 326: 379-389.

Jung K.-H., Trivedi V.D., Spudich J.L., 2003, 'Demonstration of a sensory rhodopsin in eubacteria', *Molecular Microbiology*, 47: 1513-1522.

Kell D.B., Oliver S.B., 2003, 'Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era', *BioEssays*, 26: 99-105.

Kitano H., 2002, 'Systems biology: a brief overview', *Science*, 295: 1662-1664.

Koonin E.V., 2007, 'Metagenomic sorcery and the expanding protein universe', *Nature Biotechnology*, 25: 540-542.

Kowalchuk G.A., Speksnijder A.G.C.L., Zhang, K., et al., 2007, 'Finding the needles in the metagenome haystack', *Microbial Ecology*, 53: 475-485.

Lanyi J.K. 2004, 'Bacteriorhodopsin', *Annual Review of Physiology*, 66: 665-688.

Lewin R., 1986, 'Proposal to sequence the human genome stirs debate', *Science*, 232: 1598-1600.

Lockhart D.J., Winzeler E.A., 2000, 'Genomics, gene expression and DNA arrays', *Nature*, 405: 827-836.

Man D., Wang W., Sabehi G., et al., 2003, 'Diversification and spectral tuning in marine proteorhodopsins', *EMBO Journal*, 22: 1725-1731.

Man-Aharonovich D., Sabehi G., Sineshchekov O.A., et al., 2004, 'Characterization of RS29, a blue-green proteorhodopsin variant from the Red Sea', *Photochemical and Photobiological Science*, 3: 459-462.

Martinez A., Bradley A.S., Waldbauer J.R., et al. 2007, 'Proteorhodopsin photosystem gene expression enables photophosphorylation in heterologous host', *Proceedings of the National Academy of Sciences*, 104: 5590-5595.

McCarren J., DeLong E.F., 2007, 'Proteorhodopsin photosystem gene clusters exhibit co-evolutionary trends and shared ancestry among diverse marine microbial phyla', *Environmental Microbiology*, 9: 846-858.

Mongodin E.F., Nelson K.E., Daugherty S., et al., 2005, 'The genome of *Salinibacter ruber*: convergence and gene exchange among hyperhalophilic bacteria and archaea', *Proceedings of the National Academy of Sciences*, 102: 18147-18152.

Morris R.M., Rappé M.S., Connon S.A., et al., 2002, 'SAR11 clade dominates ocean surface bacterioplankton communities', *Nature*, 420 : 806-810.

Oesterhelt D., Stoeckenius W., 1971, 'Rhodopsin-like protein from the purple membrane of Halobacterium halobium', *Nature New Biology*, 233: 149-152.

Oesterhelt D., Stoeckenius W., 1973, 'Functions of a new photoreceptor membrane', *Proceedings of the National Academy of Sciences*, 70: 2853-2857.

Oremland R.S., Capone D.G., Stolz J.F., Furhman J., 2005, 'Whither and wither geomicrobiology in the era of 'community genomics'?', *Nature Reviews Microbiology*, 3: 572-578.

Pace N.R., 1997, 'A molecular view of microbial diversity and the biosphere', *Science*, 276: 734-740.

Patterson D.J., and Sogin, M.L., 2000. Eukaryotes. Eukaryota, organisms with nucleated cells. *Tree of Life Web Project*
www.tolweb.org/Eukaryotes/3/2000.09.08 (accessed March 25th, 2008).

Pazos F., Valencia A., De Lorenzo V., 2003, 'The organization of the microbial biodegradation network from a systems biology perspective', *EMBO Reports*, 4, 994-99.

Ram R.J., VerBerkmoes N.C., Thelen M.P., et al., 2005, 'Community proteomics of a natural microbial biofilm', *Science*, 308, 1915-1920.

Rheinberger H.-J., 1997, *Toward a history of epistemic things: synthesizing proteins in the test tube*, Stanford, CA: Stanford University Press.

Rheinberger H.-J., McLaughlin P., 1984, 'Darwin's experimental natural history', *Journal of the History of Biology*, 17: 345-368.

Ribe N., Steinle F., 2002, 'Exploratory experimentation: Goethe, land, and color theory', *Physics Today*, 55 (7): e43.

Riesenfeld C.S., Schloss P.D., Handelsman J., 2004, 'Metagenomics: genomic analysis of microbial communities', *Annual Reviews of Genetics*, 38, 525-52.

Ruiz-González M.X., Márin I., 2004, 'New insights into the evolutionary history of Type I rhodopsins', *Journal of Molecular Evolution*, 58: 348-358.

Rusch D.B., Halpern A.L., Sutton G., et al., 2007, 'The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through Eastern Tropical Pacific', *PLoS Biology*, 5 (3): e77.

Sabehi G., Béjà O., Suzuki M.T., et al., 2004, 'Different SAR86 subgroups harbour divergent proteorhodopsins', *Environmental Microbiology*, 6: 903-910.

Sabehi G., Loy A., Jung K.-H., et al., 2005, 'New insights into metabolic properties of marine bacteria encoding proteorhodopsins', *PLoS Biology*, 3 (8): e273.

Sharma A.K., Spudich J.L., Doolittle W.F., 2006, 'Microbial rhodopsins: functional versatility and genetic mobility', *Trends in Microbiology*, 14: 463-469.

Sharma A.K., Zhaxybayeva O., Papke R.T., Doolittle W.F., 2008, 'Actinorhodopsins: proteorhodopsin-like gene sequences found predominantly in non-marine environments', *Environmental Microbiology*, 10: 1039-1056.

Spudich J.L., 2006, 'The multitalented microbial sensory rhodopsins', *Trends in Microbiology*, 14: 480-487.

Spudich J.L., Yan C.-S., Jung K.-H., Spudich E.N., 2000, 'Retinylidene proteins: structures and functions from archaea to humans', *Annual Review of Cell and Developmental Biology*, 16: 365-392.

Sreenivasan A., 2001, 'Earth's anonymous inhabitants', *Science Notes*, UCSC Science Communication Program.
Scicom.ucsc.edu/SciNotes/0101/bacteria.html

Staley J.T., Konopka A., 1985, 'Measurements of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats', *Annual Review of Microbiology*, 39: 321-346.

Steinle F., 1997, 'Entering new fields: exploratory uses of experimentation', *Philosophy of Science*, 64: S65-S74.

Steinle F., 2002, 'Experiments in history and philosophy of science', *Perspectives on Science*, 10: 408-432.

Stevenson B.S., Eichorst S.A., Wertz J.T., et al., 2004, 'New strategies for cultivation and detection of previously uncultured microbes', *Applied and Environmental Microbiology*, 70: 4748-4755.

Steward G.F., Rappé M.S., 2007, 'What's the 'meta' with metagenomics?', *ISME Journal*, 1: 100-102.

Thomas C.M., and Nielsen K. M., 2005, 'Mechanisms of, and barriers to, horizontal gene transfer between bacteria', *Nature Reviews Microbiology*, 3: 711-721.

de la Torre J., Christianson L.M., Béjà O., et al., 2003, 'Proteorhodopsin genes are distributed among divergent marine bacterial taxa', *Proceedings of the National Academy of Sciences*, 100: 12830-12835.

Tringe S.G., Rubin, E.M,. 2005, 'Metagenomics: DNA sequencing of environmental samples', *Nature Reviews Genetics*, 6: 805-814.

Tyson G.W., Chapman J., Hugenholz P., et al., 2004, 'Community structure and metabolism through reconstruction of microbial genomes from the environment', *Nature*, 428: 37-43.

Venter J.C., Remington K., Heidelberg J.F., et al., 2004, 'Environmental genome shotgun sequencing of the Sargasso Sea', *Science*, 304, 66-74.

Walter J.M., Greenfield D., Bustamante C., Liphardt J., 2007, 'Light-powering *Escherichia coli* with proteorhodopsin', *Proceedings of the National Academy of Sciences*, 104: 2408-2412.

Wang W.-W., Sineschekov O.A., Spudich E.N., Spudich J.L., 2003, 'Spectroscopic and photochemical characterization of a deep ocean proteorhodopsin', *Journal of Biological Chemistry*, 278: 33985-33991.

Ward N., 2006, 'New directions and interactions in metagenomics research', *FEMS Microbiology Ecology*, 55: 331-338.

Waschuk S.A., Bezerra Jr. A.G., Shi L., Brown L. S., 2005, '*Leptosphaeria* rhodopsin: bacteriorhodopsin-like proton pump from a eukaryote', *Proceedings of the National Academy of Sciences*, 102: 6879-6883.

Waters C. K., 2004, 'What was classical genetics?', *Studies in History and Philosophy of Science*, 35: 783-809.

Wilmes P., Bond P.L., 2006, 'Metaproteomics: studying functional gene expression in microbial ecosystems', *Trends in Microbiology*, 14, 92-97.

Woese C.R., Fox G.G., 1977, 'Phylogenetic structure of the prokaryotic domain: the primary kingdoms', *Proceedings of the National Academy of Sciences*, 11: 5088-5090.